



### SciDAC Meeting to Showcase LBNL's Leadership in Projects

Since DOE launched its Scientific Discovery through Advanced Computing (SciDAC) program in 2001, scientists at Berkeley Lab have led a number of the projects and provided strong support for many others.

When the annual SciDAC meeting is held June 26–30 in San Francisco, LBNL's contributions to the success of the program will be highlighted via invited talks, panel discussions and poster presentations. Here is a rundown of Berkeley Lab contributions to the meeting program.

#### Invited talks:

- “Spatial Modeling in Systems Biology,” Phil Colella, CRD
- “New Approaches to Fast Electron Correlation Methods,” Martin Head-Gordon, Chemical Sciences Division/UC Berkeley
- “Numerical Simulation of Laboratory-Scale Premixed Turbulent Flames,” Marc Day, CRD
- “The Future of Numerical Linear Algebra Libraries, Automatic Tuning of Sparse Matrix Codes, the Next LAPACK and ScALAPACK,” Jim Demmel, CRD/UC Berkeley
- “Solving Large-Scale Eigenvalue Problems in SciDAC Applications,” Chao Yang, CRD

#### Poster presentations:

- “SciDAC Advances and Applications in Computational Beam Dynamics,” Robert

(continued on page 2)

### CRD Report

CRD Report is published every other month, highlighting recent achievements by staff members in Berkeley Lab's Computational Research Division. Distributed via email and posted on the Web at <http://crd.lbl.gov/DOEResources>, CRD Report may be freely distributed. CRD Report is edited by Jon Bashor, [JBashor@lbl.gov](mailto:JBashor@lbl.gov) or 510-486-5849.

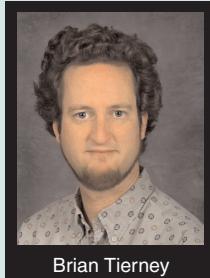
### NetLogger Helps Supernova Factory Improve Data Analysis

The Nearby Supernova Factory (SNfactory) project, established at Berkeley Lab in 2002, aims to dramatically increase the discovery of nearby Type 1a supernovae by applying assembly-line efficiencies to the collection, analysis and retrieval of large amounts of astronomical data.

To date, the program has resulted in the discovery of about 150 Type 1a supernovae — about three times the entire number reported before the project was started. Type Ia supernovae are important celestial bodies because they are used as “standard candles” for gauging the expansion of the universe.

Contributing to the SNfactory's remarkable discovery rate is its custom-developed “data pipeline” software. The pipeline fills with up to 50 gigabytes (billion bytes) of data per night from wide-field cameras built and operated by the Jet Propulsion Laboratory's Near Earth Asteroid Tracking program (NEAT). NEAT uses remote telescopes in Southern California and Hawaii.

Around 25,000 new images are captured each day, and the goal is to complete all processing before the next day's images arrive. Image data is copied in real time from the Mt. Palomar Observatory in Southern California to a mass storage system at NERSC. Then



Brian Tierney



Daniel Gunter

the image data is copied to a large shared disk array on a 344-node cluster called PDSF. Each image is 8 MB (uncompressed), and the processing of each image requires between 5 and 25 reference images, for a total disk space requirement of about 0.5 TB each day.

Supernovae are found by comparing recently acquired telescope images with older reference images. If there is a source of light in the new image that did not exist in the old image, it could be a supernova. Subtracting the new image from the reference image identifies new light sources. This process is quite delicate: aligning the images, matching the point-spread functions, and matching the photometry and bias all require precise calibration.

(continued on page 2)

### DSD Staff to Figure Prominently at HPDC 14

When the Fourteenth IEEE International Symposium on High-Performance Distributed Computing convenes in July, Berkeley Lab's Distributed Systems Department (DSD) will have a strong presence.

Not only are two of the papers and one of the posters from LBNL, but Deb Agarwal and Mary Thompson are also members of the program committee. Co-sponsored by IEEE and ACM, HPDC-14 will be held July 24–27 in Research Triangle Park, North Carolina.

The 14th annual meeting of the High Performance Distributed Computing conference series provides a forum for presenting the latest research findings on the design and use of parallel and distributed systems for high end computing, collaboration, data analysis and other innovative applications.

Here is a look at the program components contributed by DSD staff:

- “State and Events for Web Services: A Comparison of Five WS-Resource Framework and WS-Notification Implementations,” paper by Marty Humphrey, Glenn Wasson, Keith Jackson, Joshua Boverhof and Matt Rodriguez

- “Simplifying FusionGrid Security,” paper by Justin Burruss, Thomas Fredian and Mary Thompson

- “Techniques for Tuning Workflows in Cluster Environments,” poster by Brian Tierney and Daniel Gunter

## NetLogger Debugs SNfactory Workflow

(continued from page 1)

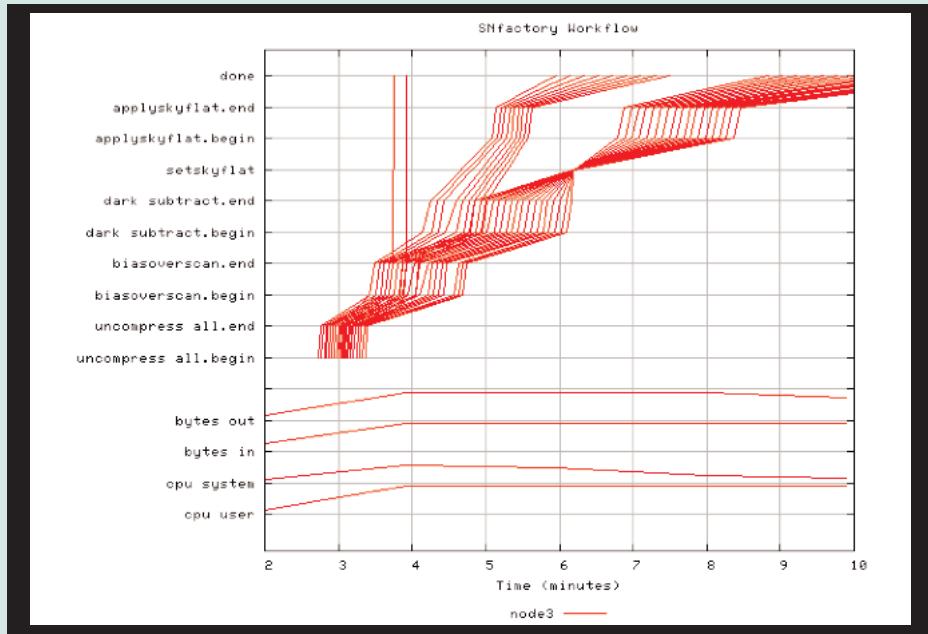
Because of the high demand put on all the resources in the pipeline, making sure that the data flow smoothly and can be analyzed quickly and correctly is critical to the overall success. While there are a number of tools for evaluating the performance of single systems, identifying the workflow bottlenecks in a distributed system such as the SNfactory requires a different type of application.

For the past 10 years, Brian Tierney and others in the Collaborative Computing Technologies Group have been developing the Netlogger toolkit as part of the Distributed Monitoring Framework project. NetLogger is a set of libraries and tools to support end-to-end monitoring of distributed applications. During the past few months, the team has been working closely with the SNfactory project to help debug and tune their application.

"NetLogger has been extremely useful in the debugging and commissioning of our data processing pipeline," said Stephen Bailey, one of the lead developers on the SNfactory project. "It has helped us identify bugs and processing bottlenecks in order to improve our efficiency and data quality. It additionally has allowed real time monitoring of the data processing to quickly identify problems that need immediate attention. This debugging, commissioning, and monitoring would have taken much longer without NetLogger."

Tierney and Bailey, along with Dan Gunter of the Collaborative Computing Technologies Group, have written a paper entitled "Scalable Analysis of Distributed Workflow Traces," which will be presented at the 2005 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'05) to be held June 27-30 in Las Vegas. The paper can be found at <<http://dsd.lbl.gov/publications/NetLogger-SNFactory.pdf>>.

"The first problem the SNfactory scientists asked us to solve was to figure out why some of their workflows were failing without any error messages as to the cause," Tierney said. "Even when error messages were generated, the SNfactory application produced thousands of log files, and it was very difficult to locate the log messages related to failed workflows. NetLogger was very useful for easily characterizing where the failures were



occurring so they would know where to focus debugging efforts."

The figure above shows a typical workflow for the SNfactory application on a single cluster node. CPU and network data is shown at the bottom. This figure actually demonstrates a bug in the SNfactory processing that went undetected for several months before NetLogger analysis.

The SNfactory application processes a group of images together, starting with uncompressing the images, and then doing image calibration and subtraction. The next step is

to generate a skyflat image, which is a calibration image that is formed from a median combination of several of other images. The skyflat is used to correct other images to adjust for the sky brightness on a given night, which can vary due to humidity, cloud cover, and so on. The skyflat calibration image is then applied to all images within the job. Under some conditions it was determined erroneously that the skyflat calibration was not necessary. All lifelines in the figure, except the two nearly vertical ones near the beginning, should have converged at the set-skyflat event.

## SciDAC Meeting Features LBNL Expertise

(continued from page 1)

Ryne, Accelerator and Fusion Research Division

- "Cartesian Grid Embedded Boundary Methods for Problems with Complex Geometries," Daniel Graves, CRD
- "The Roles of Sparse Direct Methods in Large-Scale Simulations," Xiaoye "Sherry" Li, CRD
- "Adaptive Mesh Refinement for Particle-in-Cell Methods," David Serafini, CRD
- "An Efficient Indexing Technology for

Accelerating Data Intensive Science," Kesheng "John" Wu, CRD

### Panel presentations:

- Rob Ryne of LBNL's Accelerator and Fusion Research Division will present the High Energy/Nuclear Physics perspective in a panel discussing "SciDAC II: The Shape of Things to Come."
- Lenny Oliker of CRD will discuss "The NERSC Benchmark Suite" in a panel session looking at "SciDAC II: Perspectives on Hardware and Software Infrastructure."

### DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California. Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.